DOCUMENT RESUME

ED 190 654 TM 800 451

AUTHOR Griffin, Andrew H., Jr.
TITLE Standardized Testing: Implications for Minority Students.
INSTITUTION National Education Association, Washington, D.C.
PUB DATE [78]
NOTE 17p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Blacks; Examiners; *Minority Groups; *Standardized Tests; *Test Bias; Testing Problems; Test Reliability; *Test Results; Test Validity
IDENTIFIERS Examiner Effect; Test Score Decline

ABSTRACT

The implications of standardized testing for minority students are explored. Test terminology is described in terms of objectivity, standardization, reliability, and validity. Primarily, however, the paper reviews the objectivity of standardized testing, that is, of those tests which are either norm-referenced or criterion-referenced. The use of these tests is illustrated. Reasons for the decline in satisfactory test results is cited, and the factors reported by the National Academy of Education Committee on Testing and Basic Skills are emphasized: (1) proliferation of courses; (2) confusion about the appropriate role of teachers; (3) slackening of "on task" attention; and (4) dismantling of opportunities for intensive study in selective academic environments at the secondary level. Other researchers are broadly cited pertaining to a critical overview of standardized testing and to some alternatives to standardized testing. (GK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
A. Griffith
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

APR 1 ?

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED190654

STANDARDIZED TESTING: IMPLICATIONS FOR MINORITY STUDENTS

By Dr. Andrew H. Griffin, Jr.

What do you do when you find out you have to take a test? Some us us immediately begin to perspire, others catch their breath, some of us want to run, some get charged up, and some remain cool. The thought of taking a test makes us think about how knowledgeable we are about the subject of the test. And for most of us the results of the test are clear: We were either right or wrong; it was our fault that we didn't answer a question correctly.

If you think that, you're partly right and partly wrong. It is the intention of this paper to explore standardized tests and their implications for minority students, hoping that when you take your next test (or give your next test) you will not so readily accept the test results or their interpretation.

Terminology

A test can be objective, proper, standardized, reliable, and valid--and still be a very bad test.[1] The four main words that resound within in the field of testing are objectivity, standardization, reliability, and validity. This paper will review primarily the objectivity of standardized testing, but defining all four terms will establish a common frame of reference.

1. Objectivity means that everyone takes a test under more or less the same conditions and all the tests are graded under more or less the same conditions.

2. Standardization means the making of arrangements for all students to take the test under similar conditions, for example, the same available time. It also means the establishment of norms for performance so that test scores come out as percentiles. Standardization has no bearing on the quality of the test; it affects only the reporting of scores. Further, standardization applies to the sample

TM 800451

population to whom the test taken is compared. The test maker has established expected standards of performance determined by administration of the test to a selected group of students, for example, by age or grade.

3. Reliability means how well a test agrees with itself. The higher the reliability the better the test. The ratings of the test are generally through the reliability coefficient. The closer to 1.0, the more reliable the test, i.e., .95 is excellent, .90 is pretty good, and .80 is not so good.

4. Validity means the degree to which a test measures what it is supposed to measure. Validity is extremely difficult to ascertain. It is determined by an expert or by comparing the test results with some other measure. If a test seems to an expert to measure what it purports to measure, then it has content or face validity. For example, it would be relatively easy to establish content or face validity of a math test designed to measure students' abilities to add one digit numbers below 10.

Comparison Validity is determined by comparing the test scores with a second measure of competence. If students score high on a test in math and receive high grades in the course, then there is a possibility that the test has a high degree of validity. The only catch is that competence in the math course should be compared to a measure other than the test.

These four terms--objectivity, standardization, reliability, and validity--should be examined each time a test is approached or presented.

Standardized tests may be of two kinds: norm-referenced or criterion-referenced. Norm-referenced tests are used to determine how the performance of a given student or group of students compares with the performance of a group of students whose scores are given as the norm. Criterion-referenced tests are used to determine whether a given student has reached a particular level of performance; they do not compare the student with other students.[2]

There has been a lot of controversy in recent years about standardized testing, whose history goes back to the success of the Army Alpha test in World War I. This paper-and-pencil test was designed to select people who would make "good" soldiers, and it also effectively screened out Blacks based upon IQ. The schools and colleges quickly adopted psychological tests to select students, and by 1929, more than 5 million tests were being administered annually. In 1975 the National Education Association reported that at least 200 million achievement test forms were being used each year in the United States, and this number represents only 65 percent of all educational and psychological tests that are administered.[3] Consequently, tests are used to make individual or institutional decisions, diagnosis or prognosis, research or evaluation.

A case in point: In 1912 Henry Goddard, one of the original translators, imported and translated the Binet Test. After testing a representative sample of new immigrants in 1912, he reported that 83 percent of Jews, 79 percent of Russians and 87 percent of Italians were feebleminded. It was pointed out to Congress again and again that, so far as IQ was concerned, immigrants from Southeastern Europe were genetically inferior to Nordic immigrants from Northwestern Europe. And, when Congress passed an immigration law in 1924, it embodied for the first time what was called "national origin quotas."[4]

The results of standardized tests indicate that students today are not performing as well on achievement tests as students of five years ago. The Wirtz Report (1977)[5] for the Educational Testing Service (ETS) cites forces outside the school that have caused the decline of writing skills and Scholastic Aptitude Test scores: (a) television, (b) divorce rates, (c) lower student motivation, (d) a decade of distraction, (e) increased absenteeism, and (f) the increasing number of persons taking the SAT. The National Academy of Education Committee on Testing and

Basic Skills (1977)[6] on the other hand reported that factors within school have been particularly important in causing the decline in writing skills and SAT scores. These are:

a. Proliferation of courses. With an increased number of options, students have been able to choose less demanding courses than the traditional requirements of English, mathematics, science, and social studies (especially history).

b. Confusion about the appropriate role of teachers. Teachers have been given new and often contradictory models of appropriate pedagogic behavior. But "the discovery method," "the open classroom," "individualized instruction," or "team teaching" were rarely accompanied by teacher training and professional development programs adequate for effective implementation.

c. The slackening of "on task" attention. The Academy reports some studies showing that even in relatively "good" classrooms students are "on task" for only 30 percent of the instructional day. It believes that more effective use of school time would be a significant reform.

d. The dismantling of opportunities for intensive study in selective academic environments at the secondary level. For example, programs for the academically talented have been closed or forced to change admission standards.

A reaction to the decline in test scores is the setting of minimum competency standards for elementary and secondary students. As of March 15, 1978, 13 states had taken some type of action to mandate minimum competencies. In the remaining states either legislation is pending or legislative or state board studies are under way.[7]

For Blacks and other minorities standardized tests are being used to determine whether intelligence is fixed at birth or depends upon environmental factors, as

set forth in the works of Arthur Jensen and William Shockley. In a 1972 study, Psychologist Jane Mercer found that from 50 to 300 percent more Blacks and Mexican Americans were identified as mentally retarded than could reasonably be expected from their proportion of the population.[8]

The National Teacher Examination, designed by the Educational Testing Service, is used in South Carolina to assess teachers' performance and production. This is a misuse of the test scores according to ETS, yet the policy is still in effect and has successfully trimmed the number of Black teachers--from 43 percent of the teaching force in 1953 to 29 percent in 1975.[9]

North Carolina uses the National Teacher Examination with a cutoff score of 950 to determine the salary, retention, and tenure of teachers with substantial in-service experience and the certification of prospective teachers with no experience. This minimum score disqualifies proportionately more Black persons than White.[10]

Blacks who wanted to become police officers in Washington, D.C., were given an application and a written test of verbal skills to determine their ability to be police officers. Two who were rejected took the case to court. The case was lost in the lower court but won on appeal. The appeal stated that since four times as many Blacks as Whites failed the test, such disproportionate impact sufficed to establish a constitutional violation unless the employer could demonstrate that the skills measured by the test were substantially related to job performance. The Supreme Court upheld that decision.[11]

Think back and note the many times a test was used as the role criterion to determine whether you were a success or a failure. Also note the times a test was used as the first of many criteria and if you failed you were not allowed to deal with the next criterion. Yes, tests play an important role throughout our lives. Our parents and peers rarely listen to our arguments regarding the fairness of the test; what is generally focused upon is the end result--your score.

Several prominent national organizations are so concerned about the misuse and abuse of standardized testing that they have called for a moratorium on the practice--The National Association for the Advancement of Colored People, The National Education Association, The National Association of Elementary School Principals and The American Personnel and Guidance Association. Their call for a moratorium to delay testing in reality was asking a 300 million dollars a year industry to take a vacation without pay. (The Bank of New York's Research Division reports that gross earnings for testing for the three major test publishers and one scoring company totaled $105 million in 1974. These companies are Houghton Mifflin, which publishes the Iowa Basic Skills; Harcourt Brace Jovanovich, which publishes Stanford Achievement and Metropolitan Achievement Tests; and the Westinghouse Measurement Research Center.[12] In 1961, the last year in which IBM-owned Science Research Associates filed an independent report, net sales totaled $9 million.)[13]

Testing is instricately woven into the fabric of the United States. Fred Hechinger in the New York Times of May 1, 1977[14] wrote: "Americans are a nation of score keepers. They want to know just exactly how they are doing in everything from sexual performance to their children's third-year experience. There is constant pressure by parents, school administrators, state education authorities, Congress and colleges to compare the performance of each child and each district with 'the norm.' Standardized tests are expected to do this job."

Since the majority of our children must go to school until the age of 16, the school is the next most important socializing force in our society, after the home. And since our educational philosophy is based mainly on the Anglo-Saxon ideal, standardized testing has important implications for minority and low-income students, particularly their objectivity of the tests, as pointed out in some of the literature. Flaugher (1970)[15] submits that there is sketchy but provocative evidence to indicate that the atmosphere, both physical and psychological, in which an examination is

completed can influence the quality of performance.

Sattler (1969)[16] reports that White examiners affect and at times impede the performance of Black children. Savage and Bower (1972)[17], using subtests of the Wechsler Intelligence Scale for children, found that both Black and White children performed better with testers of their own race on block design but not on digit span. Katz (1970)[18] found that the race-of-the examiner effect on Blacks is dependent upon the type of task as well as its complexity and suggests that variables such as speed and interaction between the examiner and examinee could heighten the sensitivity during the test. Hawkes and Koss (1970)[19] report that inner-city students exhibit high levels of general and test-specific anxiety. In the Savage and Bowers[20] study, the race-of-tester effect was mediated by the racial make-up of the school--monoracial White, monoracial Black, multiracial--and the age of the children. In the first grade, both Black and White children performed better with some race testers. Across school types however, the effect was stronger in the all Black school. There was less tester difference in third and fifth grades, indicating that the older children had less reaction to the White teacher. Black first-and third-graders in multiracial schools scored consistently higher under Black testers on both tasks--digit span and block design. Only at grade five was the race-of-tester effect eliminated. The authors conclude that "contrary to popular conception, interracial contact appears to have a negative effect when applied to a biracial test situation." Studies by Williams (1969)[21], Roper (1972)[22], and Cohen and Roper (1972)[23] reported that the school environment depresses the performance of Blacks. Katz, Atchision, and Epps and Perry[24] further report that the type of feedback given to Black college students in a biracial test situation will affect their performance.

Watson (1972)[25], who replicated the works of Katz in Great Britain with West Indian teenagers, found that Black children who had a White tester and who believed

they were doing an IQ test did less well than Black children who had a Black tester and who thought the test was a research tool. Upon repeating the experiment two years later with 14-and 15-year-olds, the White tester produced a significant drop in performance. However, with a White tester who gave no instruction and a Black tester who did give instructions, a small rise in performance was noted. With younger children, performance was worse with a White tester in all instruction conditions.

Thomas et al.(1971)[26] reported that Puerto Rican students tested and retested by experienced female Puerto Rican bilingual examiners produced significantly higher scores with examiners who made them feel relaxed and comfortable than with more formal examiners who carried out the test according to instructions. Five percent of the children who were tested by the less formal examiner tested in the borderline defective range, compared with 45 percent examined by the formal examiner.

Integrated Education (1972)[27] reported that the U.S. District Court temporarily enjoined a San Francisco school district from placing Black children in classrooms for the mentally retarded based upon IQ scores because Blacks were being so placed at a rate nearly 2-1/2 times their school population. The Bay Area Association of Black Psychologists, using optimizing techniques, reexamined the seven Black children who were the plaintiffs. All of them scored significantly above the cutoff point of 75. The optimizing techniques were increasing rapport, overcoming defeatist attitudes and distractability of the children.

Paretti (1975)[28] examined sex and race with 268 Black school children in 24 different fifth-and sixth-grade classrooms. He reported that on the average subjects performed significantly lower with females than with males.

Katz-Zalk (1976)[29] report the differences in response patterns were found as a function of age, race of examiner, race of subject and in some instances, gender.

Nober and Seymour (1974)[30] noted that White student teacher speech recognition

was significantly lower for Black speakers than for White speakers. While Black listeners scoring White speakers equalled Black listeners scoring Black speakers.

Green et al. (1975)[31] report that naive test takers had significant score improvements when no time limit was imposed. They further report that middle-class children are more sophisticated in taking tests than educationally "underclassed" children, and that most minority and poor children tend to be less motivated than middle-class children. Both factors can affect test performance irrespective of ability or knowledge.

Green et al. cite two basic dimensions of test-taking ability: (1) general know-how, which encompasses such strategies as how to pace oneself, how to avoid unnecessary errors, knowing when to guess, and how to choose the correct answer by eliminating incorrect items; and (2) "test-wiseness," which is the ability to take advantage of irrelevant clues in test items to help answer questions without necessarily knowing the content.

Finally, the computerized scanner capable of scoring up to forty thousand tests an hour provides a quantitative product for each test taken. This highly sensitive machine assures high volume scoring and continuity of the concept of large-scale group testing. What it does not do is allow for the test taker's knowledge; an answer is either wrong or right. The scanner doesn't know if the question was ambiguous or misread, or if the test taker just forgot the answer. It doesn't know how the person was feeling when taking the test. However, the scoring of the test was objective.

What happens with some of the test results:

1. The majority of the students are grouped by ability. (Findley and Byron, 1970, I.)[32]

2. Low achievers are grouped together and deprived of the stimulation of high achieving children as learning models and helpers. (Findley and Byron, 1970, IV.)[33]

3. Principals and school superintendents are regularly judged by and rewarded for their pupils' performance.

4. Students are admitted to or rejected by a course, college, or professional school on the basis of test results.

5. Test results determine the curriculum and how it will be taught in the class.

6. Scores are permanently attached to student records.

7. The results maintain the differential status of population subgroups. (Mercer, 1974.)[34]

8. Test results legitimize the dominance of the Anglo cultural tradition.[35]

9. They tend to calm parents who are seeking to change educational programs.[36]

10. Cities, school districts, and schools are compared and allocated resources on the basis of test results.

To expand on this point, the National Institute of Education has developed a proposal to allocate Title I funds to school districts on the basis of reading scores of 9-year-olds every three years rather than on family poverty levels. The proposal calls for $7.2 million per state[37] over a three-year period to establish a system to collect achievement data.

Alternatives

What are some of the alternatives to standardized tests?

Perrone (1975)[38] suggests interview techniques as adopted by the University of North Dakota and Prospect School in Bennington, Vermont. This technique requires systematic documentation of information about a child from the child's teacher and parents. Perrone further suggests the increased use of diagnostic and criterion-referenced tests.

Mercer (1974)[39] presents a multicultural model which requires at least the following changes:

1. Assuming that there are as many normal curves for behavior as there are

distinct life styles a multicultural classification system would be based on multiple normal curves and would not evaluate all human behavior with a single statistical distribution.

2. Since persons from different cultural backgrounds are, in fact, from statistically different populations, they would not be combined in a single aggregate for the purpose of establishing norms. The behavior of persons from one cultural tradition would not be evaluated against the behavioral norms based on persons from a different cultural heritage.

3. Multiple measures of validity would be used. The validity of a measurement technique under the Anglo-conformity model is determined by its ability to predict academic success in an Anglo-American public school system. Determining the validity of a measurement technique in predicting success in a Mexican American, Puerto Rican, Native American, or Black cultural setting would necessarily require different measurement criteria.

4. A multicultural perspective would require clear differentiation between prognosis and diagnosis, a distinction that is frequently overlooked in present monocultural evaluation.

It would appear that standardized tests have limits just as any tests have. Some of these limits are:

a. Such a test measures what a person knows, not how he/she uses what is known.
b. The test measures what a pupil can do, not what he/she typically does.
c. The test provides multiple choice and does not necessarily reflect a person's thinking.
d. The score is just for a moment in time and not one that is fixed.
e. The score measures only one dimension of the person's life and neglects aspects such as energy, ambition, determination, adroitness, likeableness, luck, perseverance, friends, resources, and politics.
f. The test is monocultural, not multicultural.

What should we do about standardized tests?

1. Find out what types of tests are being administered by the school districts.
2. Use services of the Association of Black Psychologists, NAACP, Black institutions of higher learning to ascertain the validity, reliability and standards of the test for minority students.
3. Research and establish how the results of the test are being used, e.g., ability grouping, special classes, special curriculum, special programs, tracking.
4. Examine the racial composition of the special classes, special programs, and level of tracks.
5. Identify the ranking of minorities on tests given. Meet with school board, school officials, local education association, to develop alternatives to standardized tests.
6. Share information with minority community and concerned citizens, e.g., ministry, colleges, parents, organizations and agencies.
7. Encourage associations such as, Black Psychologists to develope a manual that will identify the different standardized tests and their negative impact on minorities.
8. Encourage the National Institute of Education to provide funding for research and training of educators and community people to develope alternatives to standardized testing.
9. Publicize the misuse and abuse of standardized tests.

Educators and concerned citizens should remember the maxims of Thomas (1977)[40]: (1) know thy test, what is being tested, how to administer the test, what reliance to place on the results; (2) know thy student, understand the population and its socioculture milieu; and (3) "know thyself"--personal attitudes and the program and practices under one's control. Additionally, educators and concerned citizens

should remember these words of Jerrold R. Zacharias, a profesor at MIT[41]. "I feel emotionally toward the testing industry as I would toward any other merchants of death. I feel that way because of what they do to the kids. I'm not saying they murder every child--only 20 percent of them."

REFERENCES

1 Lazarus, Mitchell, Coming to Terms with Testing, The National Elementary Principal Volume 54, Number 6, July-August 1975

2 Brady, Elizabeth H. To Test or Not to Test, American Educator, Winter 1977

3 Houts, Paul C. Standardized Testing in America, II. The National Elementary Principal, Volume 54, Number 6, July-August 1975

4 Rather, Dan The IQ Myth, CBS News, April 22, 1975

5 Kelley, E.W. Politics of Proficiency Report to NIE - P-770223

6 National Academy of Education Committee, Improving Educational Achievement Report to Assistant Secretary of Education, Washington, DC February 22, 1978

7 Pipho, Chris Minimum Competency Testing in 1978: A Look at State Standards, Phi Delta Kappan, May 1978

8 Houts, Paul H., IQ Tests Once Again Disturbs Educators, N.Y. Times May 1, 1977

9 Jackson, Jesse L., Teacher Examinations: A New Tool for Racists, Los Angeles Times, May 6, 1978

10 Williams, Howard P., et al. U.S.A. Plaintiff and North Carolina Association of Educators, et al. Plaintiff Intervenors vs. State of North Carolina, May 1, 1975

11 Washington, Walter E., et al. Petitioners vs. Alfred E. Davis et al. respondents Supreme Court of the United States, No. 74-1492, June 1976

12 Houts, Paul L. Standardized Testing in America, II, The National Elementary Principal, Volume 54, Number 6, July-August 1975

13 Ibid. p. 13

14 Hechinger, Fred M., Why Schools Use Standardized Tests, N.Y. Times, May 1, 1977

15 Flaugher, R.L. Testing Practices, Minority Groups and Higher Education: A Review and Discussion of the Research. Research Bulletin of the Educational Testing Service, June 1970

16 Sattler, J.M., Racial Experimenter Effects in Experimentation, Testing, Interviewing, and Psychotherapy, Psychological Bulletin 1970

17 Savage J.E., Jr. and Bowers, N.D. Tester's Influence on Children's Intellectual Performance, Washington, DC U.S. Office of Education (ERIC microfiche, No. 064329) 1972

18 Katz, I., Experimental Studies of Negro-White Relationships in Advances in Experimental Social Psychology 5 ed. L. Berkowitz, New York: Academic Press, 1970

19 Hawkes, T., and R.H. Koff, Differences in Anxiety of Private School and Inner City Public Elementary School Children Psychology in the Schools, 1970

20 Savage, J.E., Jr., and Bowers, N.D. Tester's Influence on Children's Intellectual Performance, Washington, D C U.S. Office of Education (ERIC microfiche No. 064329) 1972

21 Williams, R.L., Scholastic Attidues of Southern Negro Students, Journal of Negro Education 1969

22 Roper, S.S., Race and Assertive Classroom Behavior Integrated Education: A Report on Race and Schools, 1972

23 Cohen and Roper Modification of Interracial Interaction Disability; a Modification of Status Characteristic Theory American Sociological Review, 1972

24 Katz, I., C.O. Atchison, E.G. Epps and A. Perry Factors Affecting Response to White Intellectual Standards at Two Negro Colleges, Psychological Reports 1970

25 Watson, P., IQ: The Racial Gap Psychology Today 1972

26 Thomas, A., M.E. Hertzig, I. Dryman, and P. Fernandez, Examiner Effect in IQ Testing of Puerto Rican Working Class Children American Journal of Orthopsychiatry 1971

27 Integrated Education Number 59, P. 15, 1972

28 Paretti, Joseph Examiner Sex, Examiner Race and Anxiety Effects on the Testing of Black School Children, Research and Technical Report. 1975

29 Zalk, Sue B., Katz, Phyllis A., Katz-Zalk Projective Test:a Measure of Racial Attitudes in Children American Psychological Association 1976

30 Nober, E. Harris, Seymour, Harry N., Speech Recognition Scores of White and Black Student-Teacher Listeners for Black and White First Grade Speakers National Institute of Education OEG-1-72-009 February 1974

31 Green, Robert L., Julie G. Nyquist, Robert J. Griffore, Standardized Achievement Testing: Some Implications for the Lives of Children, National Institute of Education Test Bias Conference December 1975

32 Findley, Warren G. and Bryon, Miriam M., Ability Grouping 1970:Part I Common Practices in the Use of Tests for Grouping Students in Public Schools ERIC, ED 048381, 1970

33 Findley, Warren G. and Bryon, Miriam M., Ability Grouping 1970: Part IV Conclusions and Recommendations ERIC, ED 048384, 1970

34 Mercer, Jane R., Latent Functions of Intelligence Testing in the Public Schools Chapter in The Testing of Black Students Ed. LaMar P. Miller Prentice Hall 1974 P. 81

35 Ibid. P. 83

36 Ibid. P. 85

37 National Institute of Education, Using Achievement Test Scores to Allocate Title I Funds September 30, 1977

38 Perrone, Vito Alternatives to Standardized Testing, The National Elementary Principal Volume 54, Number 6, July-August 1975

39 Mercer, Jane Latent Functions of Intelligence Testing in the Public Schools Chapter in The Testing of Black Students Ed. LaMar Miller Prentice Hall 1974 P. 92

40 Thomas, Charles Testing and the Evaluation of the Minority Child: Some Implications for Compensatory Education Evaluations Viewpoints 53 4 31-49 July 1977

41 Zacharias, Jerold Quote P. 15 The National Elementary Principal